Jour. Agric. Res. China (中華農業研究) 43(3)：283～292 (1994)

穩定性迴歸分析之 SAS 程式[1]

呂　秀　英[2]

摘要：利用直線迴歸分析法來解釋植物穩定性的問題，是本省目前最常用而有效的方法。本文以個人電腦之 SAS 軟體中最基本的 STAT 功能系統，編寫出一個穩定性迴歸分析程式，並以一數例說明其進行分析的流程及結果，以幫助農業研究工作者進行區域試驗資料之穩定性迴歸分析。程式之輸出結果內容包含有綜合變方分析、各品系在不同環境下之產量平均值、環境指標估值、迴歸分析之統計值摘要表、穩定性係數對品系平均產量之圖示結果。

關鍵詞：SAS 程式、穩定性、迴歸分析、區域試驗、環境指標。

利用直線迴歸分析法來解釋植物穩定性的問題，是目前最常用而有效的方法(3,7)。由統計上之觀點，該法雖有一些前提之限制(5,8,9)，但與其他方法比較，在實際應用上確具有操作簡單與符合生物解釋的優點(2,4)，在未來研究基因型與環境交感上仍扮演著重要的角色。迄今，本省農業育種工作仍以迴歸係數作為區域試驗穩定品種選拔的依據。

直線迴歸法應用於穩定性分析的程序，主要有兩個步驟：首先以各環境之平均收量與整個試驗的總平均產量之偏差，作爲該一試驗的環境效應估值，稱之爲環境指標（environmental index），然後將各基因型在各環境之產量隨這些指標作迴歸分析，並以最小平方法（least square method）估計迴歸係數，利用這些迴歸係數估計值可以評估基因型的穩定性。一般電腦統計套裝軟體均含有直線迴歸分析方法，但必須先求得環境指標作爲自變數之後始可進行分析，在運算上多了一道手續，無法一氣呵成，此對於非統計專家或不熟習程式撰寫的大部份農業研究工作者而言，往往造成困擾。對此，臺灣省農業試驗所農藝系試驗統計室已在 HP9000小型電腦上利用 FORTRAN 語言撰寫一個區域試驗穩定性迴歸分析程式，提供所內研究同仁使用。目前，有鑑於本省各農業研究單位均有個人電腦，且安裝有 SAS（Statistical Analysis System）套裝軟體的機會愈來愈多，如能發展出一套完整的穩定性迴歸分析之 SAS 程式，當能提高區域試驗資料分析的速度及準確度，但是大部份的農業研究工作者對 SAS 之各種程序（procedure）語言尚無法應用裕如，爲減少使用者誤用的機會，本文特以 SAS 軟體中最基本的 STAT 功能系統(10)，編寫出一個整合性的穩定性迴歸分析程式，並以實際區域試驗資料之數例說明其進行分析的流程及結果。

統計模式與分析方法

基因型（品種）i 在環境 j 的平均性狀觀測值 Y_{ij}，以直線迴歸模式表示：

$$Y_{ij} = m_i + b_i E_j + e_{ij}$$

其中，

1. 臺灣省農業試驗所　研究報告第 1729 號。
2. 本所農藝系副研究員。臺灣省　臺中縣　霧峰鄉。

$i=1\cdots\cdots k$
$j=1\cdots\cdots n$

E_j 爲環境 j 的效應（環境指標），$\hat{E}_j=\bar{Y}_{.j}-\bar{Y}_{..}$，而 $\bar{Y}_{.j}=\sum_{i=1}^{k}Y_{ij}/k$，$\bar{Y}_{..}=\sum_{i=1}^{k}\sum_{j=1}^{n}Y_{ij}/kn$；$Y_{ij}$ 隨環境指標 E_j 作迴歸，利用最小平方法可估得截距（平均產量）m_i 及迴歸係數 b_i；而 e_{ij} 爲迴歸剩餘機差項，滿足平均爲0，變方爲 σ^2 之常態分布。

當直線迴歸關係顯著存在，可利用 b_i 作爲穩定性係數（stability index）以評價各基因型對環境的適應能力。由於育種研究上欲選拔優良的品系，須兼顧穩定性及生產力，故應同時以 b_i 與 m_i 爲選拔依據，來選拔穩定性高（$b_i=1$）而表現良好（m_i 高於對照品系或總平均值）的品種，一般以圖示表達方式，即以 m_i 爲 x 軸，b_i 爲 y 軸，觀察各品種點落在圖上區間的分布情形以進行相對間的比較。

一個穩定性迴歸分析的 SAS 程式列於表 1。表中被兩個/*符號所包含於內的文字，爲程式說明，純粹供輔助讀者瞭解各程序步驟之用，SAS 並不會執行，因此在實際輸入時可除去無妨。程式內容共有以下主要的程序步驟：資料讀入，綜合變方分析之執行，總平均值之計算，每一品系在各環境下所有重複之產量平均值之計算，環境指標之估算，迴歸分析，迴歸離差之顯著性測驗，統計值摘要表之建立，以及 b_i 對 m_i 之作圖。

數　例

今舉一實例說明分析的流程及結果，資料來源爲1986至1987年之本省 5 個飼料玉米綜合品系在 3 個地區之春作籽實產量區域試驗資料[6]，該試驗方法採逢機完全區集設計，重複 4 次，依本省雜糧作物育種實施程序，一般以年度及地區之組合作爲環境變異，故合計 5 個品系 6 個環境 4 個重複進行穩定性分析，其輸入資料檔列如表 2，資料檔係以先讀重複，再讀環境，最後讀品系的輸入格式建立，因此該檔案共10行12欄的資料中，第 1 行之12個數據依次爲品系 1 在第 1 年 3 個地區之 4 個重複的觀測值，第 2 行爲品系 1 在第 2 年 3 個地區之 4 個重複的觀測值，第 3 行爲品系 2 在第 1 年 3 個地區之 4 個重複的觀測值等等，以下類推。倘欲將資料輸入格式改以先讀重複再讀品系後讀環境的方式，則必須將程式中的資料輸入迴圈（DO—loops）更改爲　DO ENV=1 TO N；DO LINE=1 TO K；DO REP=1 TO L。程式在資料開始讀入之前，首先必須宣告品系數（K）、環境數（N）及重複數（L）的值，如本例各爲 5、6 及 4，此後在該程式之所有步驟中的敍述或數值，則可毋須再作任何更動，除非使用者欲自行增減某些輸出報表的列印結果。在某些特殊場合，研究者的作物區域試驗資料可能僅有單一重複，此時如欲應用本程式進行穩定性分析，則可將重複數（L）設定爲 1，但必須注意在其輸出結果之變方分析中的機差項將無法估計，同時也因此導致統計值摘要表中的迴歸離差無法進行顯著性測驗，除此之外，穩定性迴歸分析的估算並不受影響，即仍然可以估算出環境指標、穩定性係數和品系平均產量，以及相關的圖示。

表1. SAS/STAT 穩定性迴歸分析程式

Table 1. SAS/STAT program for regression analysis of stability

```
DATA REGSTAB ;
INFILE 'A: REGSTAB. DAT' ;
K=5 ;
N=6 ;
L=4 ;
DO LINE=1 TO K ;        /*A two-way table with K lines x N environments x L replicates*/
   DO ENV=1 TO N ;
      DO REP=1 TO L ;
         INPUT YIELD @@ ;
         OUTPUT ;
      END ;
   END ;
END ;
TITLE '—— Combined Analysis of Variance ——' ;
PROC ANOVA OUTSTAT=STAT ;                          /*Performing Combined ANOVA*/
   CLASS REP ENV LINE ;
   MODEL YIELD=REP(ENV) ENV LINE LINE*ENV ;
   TEST H=ENV E=REP(ENV) ;
PROC MEANS MEAN NOPRINT DATA=REGSTAB ;             /*Calculation of grand mean*/
   VAR YIELD ;
   OUTPUT OUT=M MEAN=GM ;
PROC MEANS MEAN NOPRINT DATA=REGSTAB ;             /*Calculation replicate means*/
   VAR YIELD ;
   BY LINE ENV ;
   OUTPUT OUT=LE MEAN=LEM ;
PROC PRINT NOOBS ;
   TITLE '—— Line means in each environment (LEM) ——' ;
   VAR LINE ENV LEM ;
PROC SORT DATA=LE ;
   BY ENV ;
PROC MEANS MEAN NOPRINT ;                          /*Estimation of environmental index*/
   BY ENV ;
   VAR LEM ;
   OUTPUT OUT=E MEAN=EM ;
DATA ENVIND ;
   IF_N_=1 THEN SET M ;
   SET E ;
   EI=EM-GM ;
TITLE '—— Environmental index ——' ;
PROC PRINT NOOBS LABEL ;
   LABEL EI='INDEX' ;
   VAR ENV EI ;
DATA COMBINE ;
   MERGE LE ENVIND ;
```

```
    BY ENV ;
    DROP _TYPE_ _FREQ_ ;
PROC SORT DATA=COMBINE ;
    BY LINE ;
TITLE '—— Regression analysis of each line ——' ;
PROC REG NOPRINT OUTEST=EST ;              /*Regression of yield on environmental index*/
    BY LINE ;
    MODEL LEM=EI/SELECTION=RSQUARE MSE SS1 ;
DATA _NULL_ ;
    SET REGSTAB ;
    IF _N_=1 ;
    CALL SYMPUT('K',K) ;
    CALL SYMPUT('N',N) ;
    CALL SYMPUT('L',L) ;
DATA _NULL_ ;
    SET STAT ;
    IF _N_=3 THEN CALL SYMPUT('ESS', SS) ;
    IF _N_=1 ;
    CALL SYMPUT('PDF',DF) ;
    CALL SYMPUT('PSS',SS) ;
DATA OUTLIST ;
    SET EST ;
    DMSE=_MSE_ ;
    SEB=SQRT(DMSE/(&ESS/(&L*&K))) ;         /*Calculation of standard error of regression*/
    DMSEF=DMSE/(&PSS/(&PDF*&L)) ;           /*Test of deviation from regression*/
    PROB=1-PROBF(DMSEF, _EDF_,&PDF) ;
    KEEP LINE INTERCEP EI SEB DMSE DMSEF PROB ;
TITLE '—— Summary statistics of stability ——' ;
PROC PRINT NOOBS LABEL DATA=OUTLIST ;      /*List of summary statistics*/
    LABEL LINE='LINE'
        INTERCEP='MEAN YIELD (LM)'
        EI='STABILITY INDEX (B)'
        SEB='STANDARD ERROR OF B'
        DMSE='DEVIATION FROM REGRESSION (D)'
        DMSEF='F-TEST VALUE OF D'
        PROB='PROB OF D' ;
    VAR LINE INTERCEP EI SEB DMSE DMSEF PROB ;
DATA _NULL_ ;                              /*Plot of stability index and mean yield*/
    SET M ;
CALL SYMPUT('REF',GM) ;
TITLE 'Regression method for stability analysis' ;
PROC PLOT DATA=OUTLIST ;
    LABEL INTERCEP='MEAN YIELD (LM)'
        EI='STABILITY INDEX (B)' ;
    PLOT EI*INTERCEP=LINE/HREF=&REF VREF=1.0 ;
RUN ;
```

表2. 本省5個玉米綜合品系在6個環境（年度×地區）之春作產量區域試驗的資料表（4重複）

Table 2. Data table of five maize composites by six environments (year×location) under spring crop season in Taiwan (four replicates)

3679	3175	3597	2704	1875	2016	2820	1939	3418	3829	3573	3951
4071	3650	3907	4055	2384	2071	1254	646	1574	1502	1998	2147
3799	3246	3279	3625	1913	1522	1543	1660	3407	2793	2969	3424
3667	3910	3932	4080	1763	1672	3134	1177	1618	1829	1067	1102
2497	2750	3091	2267	2007	2015	2482	1481	2792	2130	2080	2496
3526	3035	2056	3062	1361	824	1125	620	979	978	1132	1543
2125	2604	2687	2452	2038	2109	2434	1800	3209	2764	1975	2350
4007	3855	3808	3302	1612	2208	1615	1948	1266	1327	1873	1489
2844	2624	2424	3118	1674	2099	1476	1796	4041	3534	3846	4230
3645	5367	4061	3642	1803	2524	1237	1994	2259	1490	1429	1534

玉米區域試驗資料進行穩定性分析之結果列於附錄1。輸出結果之內容共有五個部份：綜合變方分析、各品系在不同環境下之產量平均值、環境指標估值、迴歸分析之統計值摘要表、穩定性係數對品系平均產量之圖示結果。各品系之產量平均值及穩定性係數估值，其實即是由 SAS 之迴歸程序（PROC REG）結果中摘錄出來，程式中特別多作此整理，使能立即得到兩個選拔介量的報表列印，是爲了讓使用者可免於從一大堆迴歸分析結果中再自行抄錄重要統計值的麻煩，因此在輸出報表設計上便不再列出迴歸分析之詳細結果，而在程式之迴歸程序敍述中（表1），加入 NOPRINT 之選項（option）。但是必須注意，一旦如此，迴歸係數的顯著性測驗結果便無法得知，對此使用者可自行將程式中的 PROC REG NOPRINT OUTEST=EST；該行敍述之 NOPRINT 選項去除，即可保留 SAS 內設之 PRINT 狀態。統計值摘要表除了包含品系平均產量及穩定性係數之兩個重要統計值以外，也列出穩定性係數之標準差。由本例之分析結果得知，該玉米區域試驗之品系、環境主效應，及兩者間之交感效應皆呈極顯著，產量總平均值爲2482.03；環境指標之範圍則在−975.23與1249.88之間，環境1、3、4爲較「優」的環境，而環境2、5、6屬較「劣」的環境；穩定性分析結果，以品系1爲最高產且穩定，品系5之平均產量僅次於品系1但表現較不穩定，品系3、4雖表現特別穩定但產量甚低。

本文所編寫之穩定性迴歸分析程式，主要是以 Finlay-Wilkinson[3]之方法爲理論基礎，即以迴歸係數與品系平均表現值爲選拔依據，來選拔穩定性高（迴歸係數近於1）而表現良好（平均產量高於對照品系或總平均值）的品系。至於 Eberhart and Russell (1966)[2]之進而要求用迴歸離差均方來評估迴歸係數估值之變異性大小的主張，目前在意義及應用上仍有甚多爭議[5,8,9]，但有鑑於目前仍不乏有人仍習用迴歸離差作爲第三個選拔介量，故本文亦將之納入統計值摘要表中，以提供使用者參考利用。統計值摘要表中除了列出迴歸離差及其顯著性測驗 F 值以外，並同時求出 F 值之分布機率值，以方便使用者可直接由機率值是否小於0.05或0.01，來判斷迴歸離差爲顯著或極顯著。例如由本文玉米區域試驗結果中得知（附錄1），僅有品系5之迴歸離差均方顯著異於0（機率值小於0.05），其他品系之迴歸離差則與0無顯著差異。

由於參試材料對任何穩定性分析都會造成限制及影響，因此利用任何電腦語言或套裝軟體所編寫的穩定性迴歸分析程式，在使用上都必須注意，僅利用少數的環境是不可能計算出有效的穩定性估計值。

附錄 1. SAS程式之輸出結果內容
Appendix 1. Results from the SAS output

— Combined Analysis of Variance —

Analysis of Variance Procedure
Class Level Information

Class	Levels	Values
REP	4	1 2 3 4
ENV	6	1 2 3 4 5 6
LINE	5	1 2 3 4 5

Number of observations in data set = 120

— Combined Analysis of Variance —

Analysis of Variance Procedure

Dependent Variable: YIELD

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	47	103951666.8	2211737.6	12.33	0.0001
Error	72	12912246.1	179336.8		
Corrected Total	119	116863912.9			

R-Square	C.V.	Root MSE	YIELD Mean
0.889510	17.06194	423.4817	2482.02500

— Combined Analysis of Variance —

Analysis of Variance Procedure

Dependent Variable: YIELD

Source	DF	Anova SS	Mean Square	F Value	Pr > F
REP(ENV)	18	2897820.65	160990.04	0.90	0.5830
ENV	5	82817888.67	16563577.73	92.36	0.0001
LINE	4	8573084.47	2143271.12	11.95	0.0001
ENV*LINE	20	9662873.03	483143.65	2.69	0.0011

Tests of Hypotheses using the Anova MS for REP(ENV) as an error term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
ENV	5	82817888.67	16563577.73	102.89	0.0001

— Line means in each environment (LEM) —

LINE	ENV	LEM
1	1	3288.75
1	2	2162.50
1	3	3692.75
1	4	3920.75
1	5	1588.75
1	6	1805.25
2	1	3487.25
2	2	1659.50
2	3	3148.25
2	4	3897.25
2	5	1936.50
2	6	1404.00
3	1	2651.25
3	2	1996.25
3	3	2374.50
3	4	2919.75
3	5	982.50
3	6	1158.00
4	1	2467.00
4	2	2095.25
4	3	2574.50
4	4	3743.00
4	5	1845.75
4	6	1488.75
5	1	2752.50
5	2	1761.25
5	3	3912.75
5	4	4178.75
5	5	1889.50
5	6	1678.00

— Environmental index —

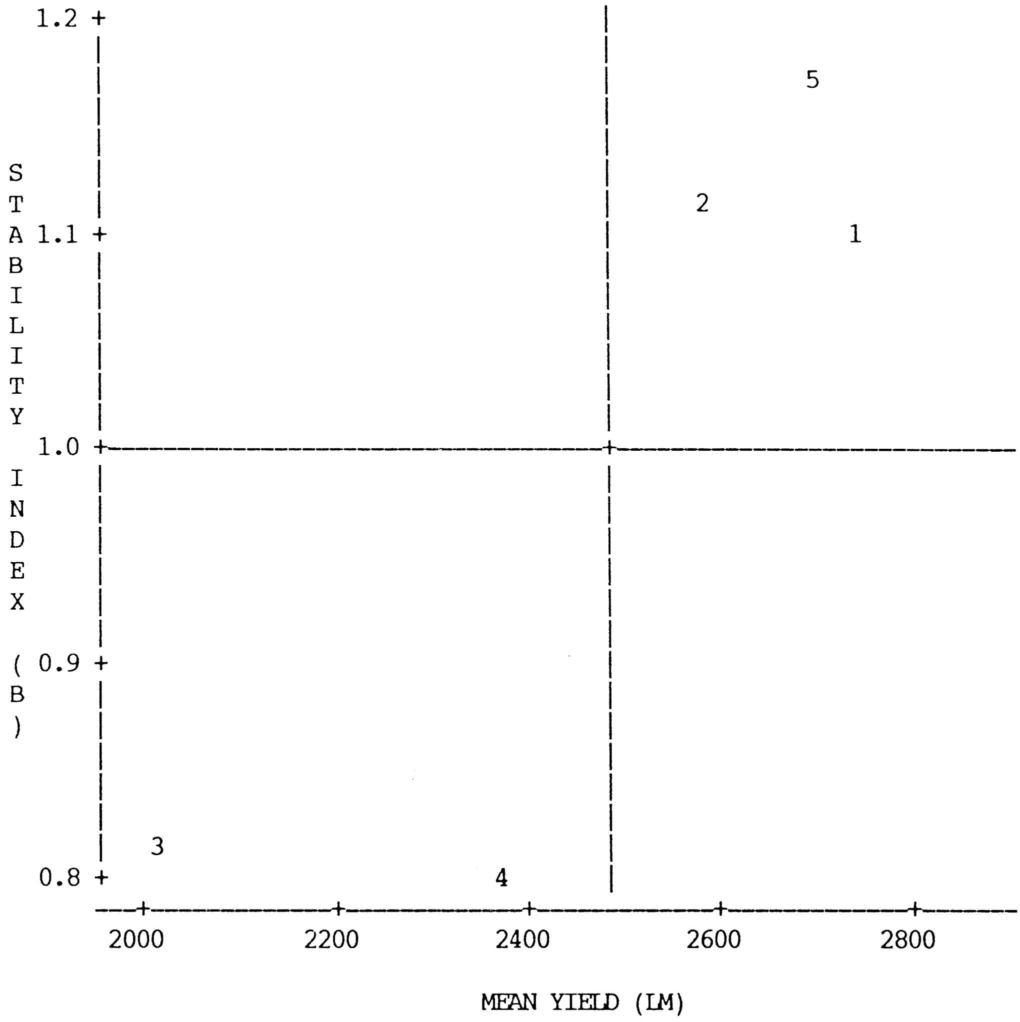
ENV	INDEX
1	447.32
2	-547.08
3	658.53
4	1249.88
5	-833.43
6	-975.23

— Summary statistics of stability —

LINE	MEAN YIELD (LM)	STABILITY INDEX (B)	STANDARD ERROR OF B	DEVIATION FROM REGRESSION (D)	F-TEST VALUE OF D	PROB OF D
1	2743.13	1.09534	0.10057	41878.52	0.93408	0.44920
2	2588.79	1.11406	0.15260	96428.69	2.15078	0.08325
3	2013.71	0.81111	0.16097	107299.21	2.39325	0.05839
4	2369.04	0.80678	0.14906	92001.04	2.05203	0.09610
5	2695.46	1.17272	0.18188	136976.48	3.05518	0.02203

Regression method for stability analysis

Plot of EI*INTERCEP. Symbol is value of LINE.



引用文獻

1. Becker, H. C. and J. León. 1988. Stability analysis in plant breeding. Plant Breed. 101:1-23.
2. Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6:36-40.
3. Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding programme. Aust. J. Agric. Res. 14:742-754.
4. Hill, J. 1975. Genotype-environment interactions—a challenge for plant breeding. J. Agric. Sci. 85:477-493.
5. Lin C. S., M. R. Binns, and L. P. Lefkovitch. 1986. Stability analysis: where do we stand? Crop Sci. 26: 894-900.
6. Lu, H. Y. and C. T. Tien. 1993. Studies on nonparametric method of phenotypic stability. II. Selection for stability of agronomic concept. J. Agric. Assoc. China 164:1-16. (in Chinese with English summary.)
7. Perkins, J. M. and J. L. Jinks. 1968. Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity 23:339-356.
8. Pritts, M. and J. Luby. 1990. Stability indices for horticultural crops. HortScience 25:740-745
9. Westcott, B. 1986. Some methods of analysing genotype-environment interaction. Heredity 56:243-253
10. SAS Institute Inc. 1990. SAS／STAT User's Guide. Version 6, 4th ed., Cary, NC.

A SAS Program for Regression Analysis of Stability[1]

Hsiu-Ying Lu[2]

Summary

Linear regression method is widely used to investigate the plant stability in Taiwan. A program written in SAS/STAT software is to presented for regressinon analysis of stability. This program runs on personal computers. The procedure is also illustrated through an example. It is useful to agricultural researchers who are analyzing the varietal performance in regional trials by regression approach. Output contains combined analysis of variance, mean yield of each variety under different environments, environmental index, summary statistics of regression analysis, plotting of stability index to varietal mean.

Key words : SAS program, Stability, Regression analysis, Regional yield trial, Environmental index.

---

1. Contribution No. 1729 from Taiwan Agricultural Research Institute.
2. Statistician, Department of Agronomy, TARI, Wufeng, Taichung, Taiwan, ROC.